Computational Visual Cognition:
Predicting Image Memorability



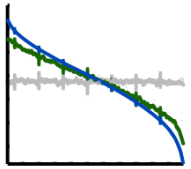

Aude Oliva

Computer Science and Artificial Intelligence Lab (CSAIL)
Massachusetts Institute of Technology
oliva@csail.mit.edu

Predicting Image Memorability

What could we do with such knowledge and technology?

Understand human memory



Diagnose memory problems

Design mnemonic aids

Advertising

Mobile Applications

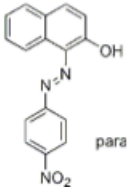
Retrieve better images from search

Logos

Social Networking

Make an image more memorable

Education



Face Memorability

Summarize photo album or video

Nature of visual long term representations

What we know …

Standing (1973)

10,000 images

83% Recognition

… people can remember thousands of images

What we don't know …

… what people are remembering for each item?

According to Standing

"Basically, my recollection is that we just separated the pictures into distinct thematic categories: e.g. cars, animals, single-person, 2-people, plants, etc.) Only a few slides were selected which fell into each category, and they were visually distinct."

"Abstract Only

Sparse Details

Highly Detailed

Massive Memory Experiment I

A stream of objects will be presented on the screen for ~ 3 second each.

Your primary task:

Remember them ALL!

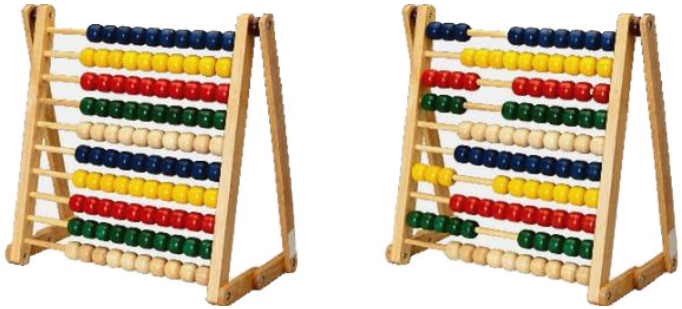
afterwards you will be tested with…

Completely different objects...

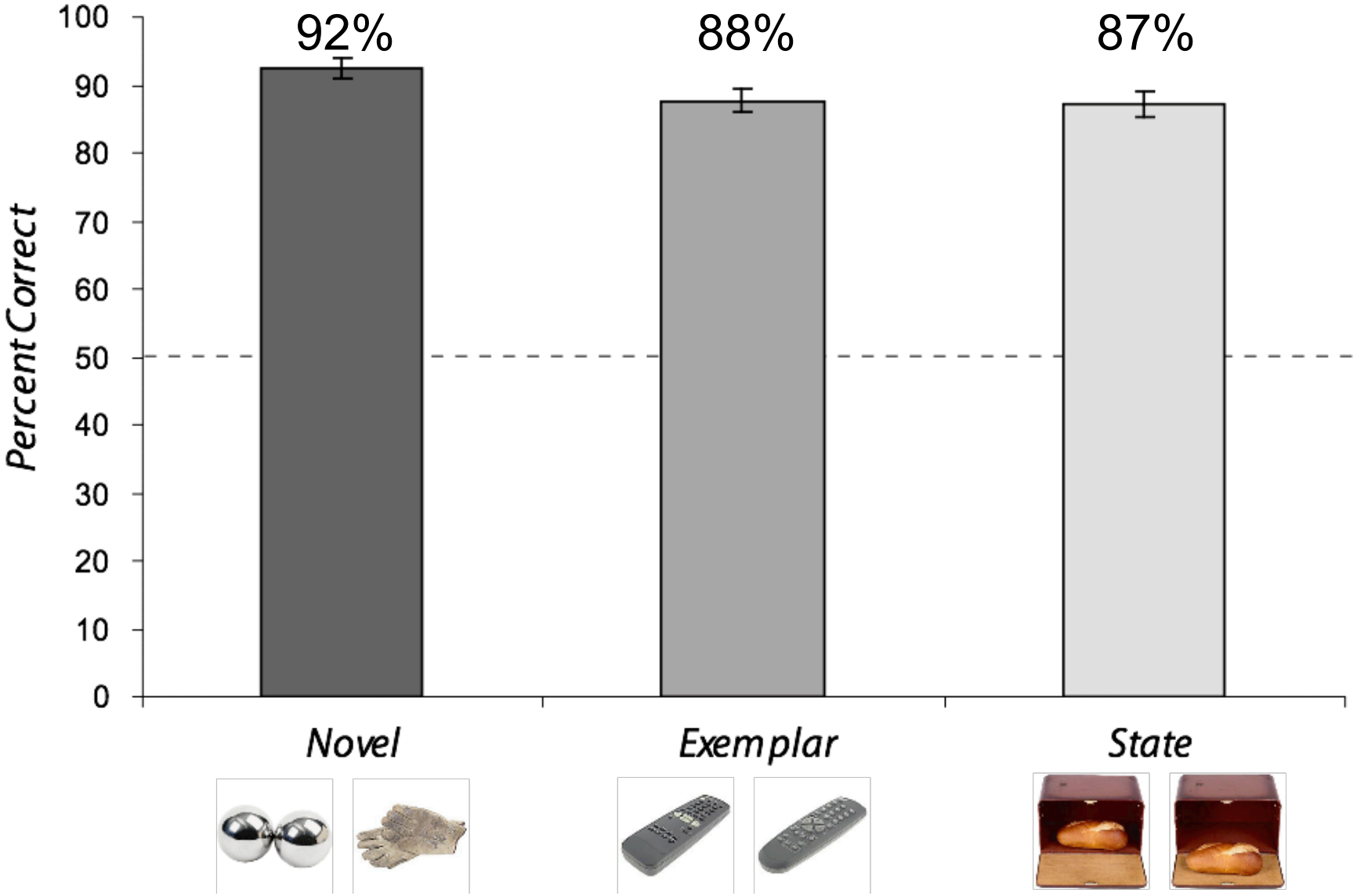
Different exemplars of the same kind of object...

Different states of the same object...

Same object, different states



Recognition Memory Results



Brady, T.F., Konkle, T., Alvarez, G.A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. Proceedings of the National Academy of Sciences, USA, vol 105 (38), 14325-14329.

Examples of **Exemplar** Memory Tests

Examples of **State** memory test


















Welcome to the

Visual Memory Game

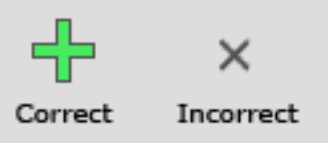
A stream of images will be presented on the screen for 1 second each.

Your task:

Clap your hands (press a key) anytime you see an image you saw before.

Be attentive, repeats may be separated by many images !

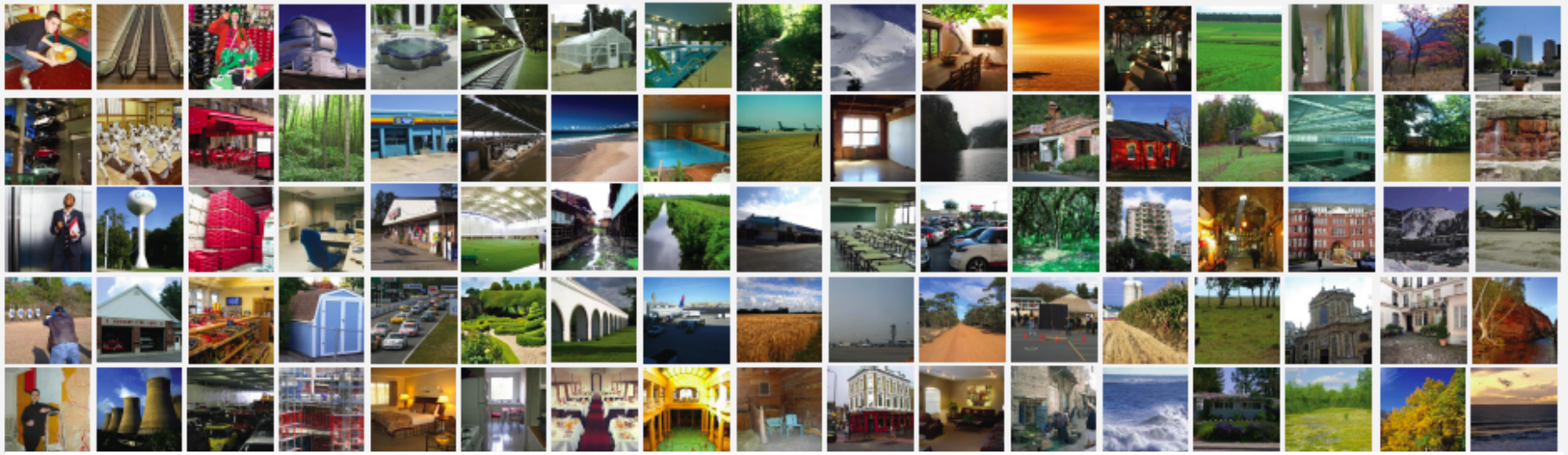
Whenever you press a key, you will get feedback:

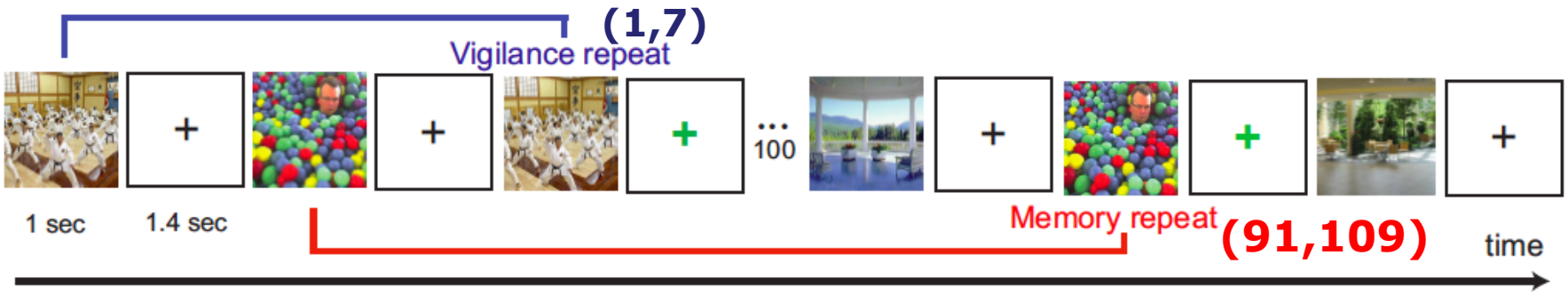


You may exit the game at any time and you will be paid in proportion to your progress at that time

Start Game!

Visual Memory Game: Method



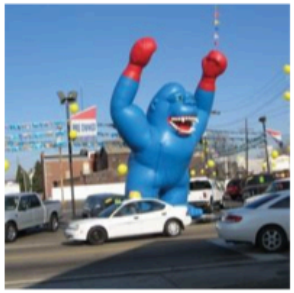


- Continuous repeat detection task
- **~ 10,000** unique images sampled from 900 scene categories (Standing, 1973; Brady et al., 2008)
- **2222** target images (memory repeats) whose repeats occurred ~ 91-109 after the first presentation

- Vigilance repeats every 1-7 images
- Each game level has 120 images
- N= 650 AMT workers
- ~ 80 scores per target images

Isola et al (2011). *IEEE Proc. Computer Vision & Pattern Recognition (CVPR)*




Memorable
Hit rate: 67/70
False alarm rate: 4/80





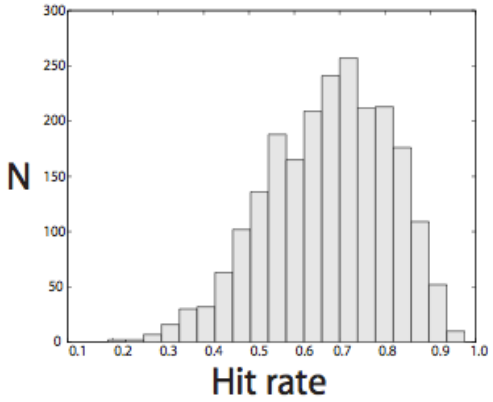
Average
Hit rate: 59/81
False alarm rate: 7/92





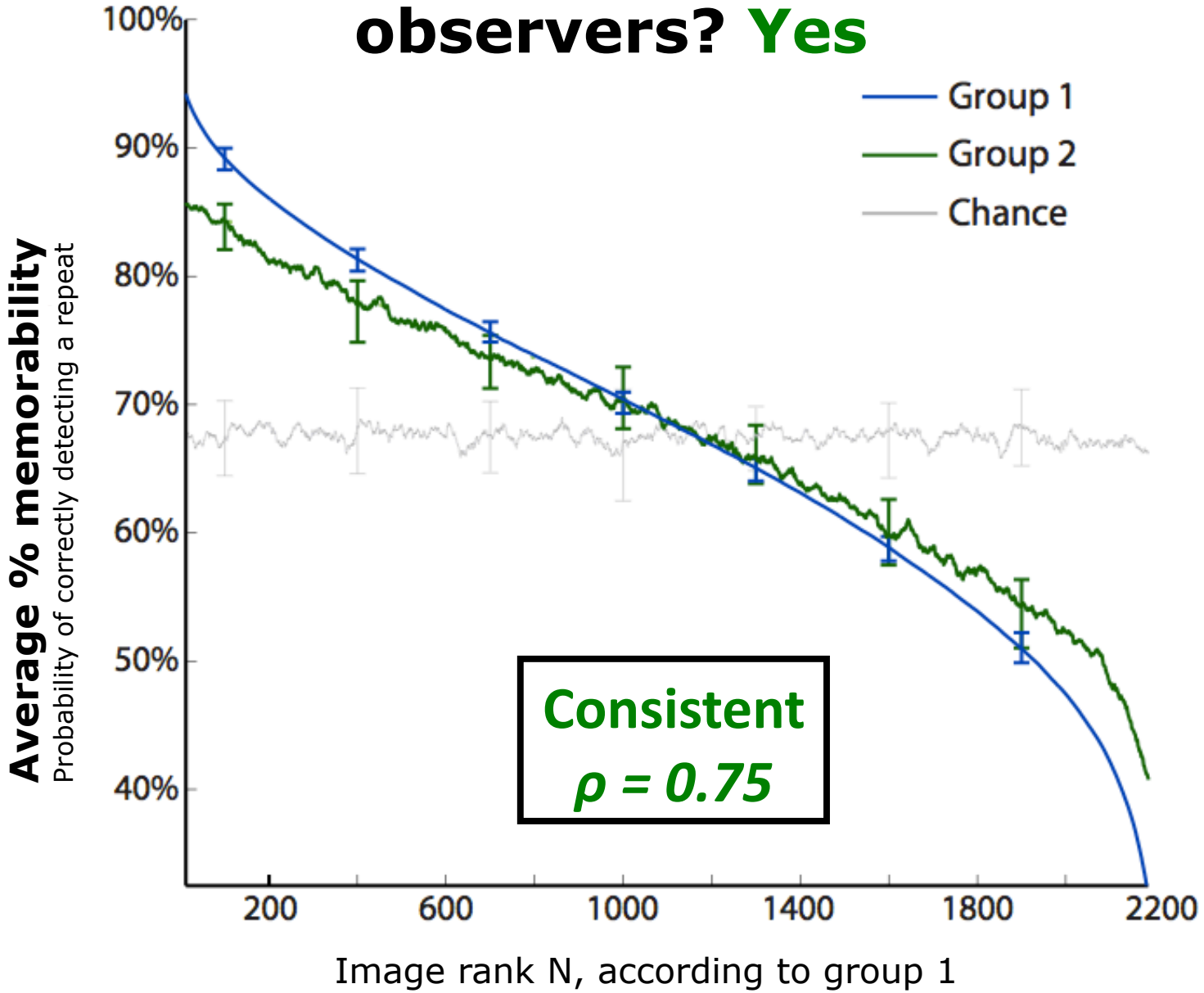
Forgettable
Hit rate: 21/68
False alarm rate: 3/82

Memorability
Mean: 67.5%
SD: 13.5%



Large differences between images

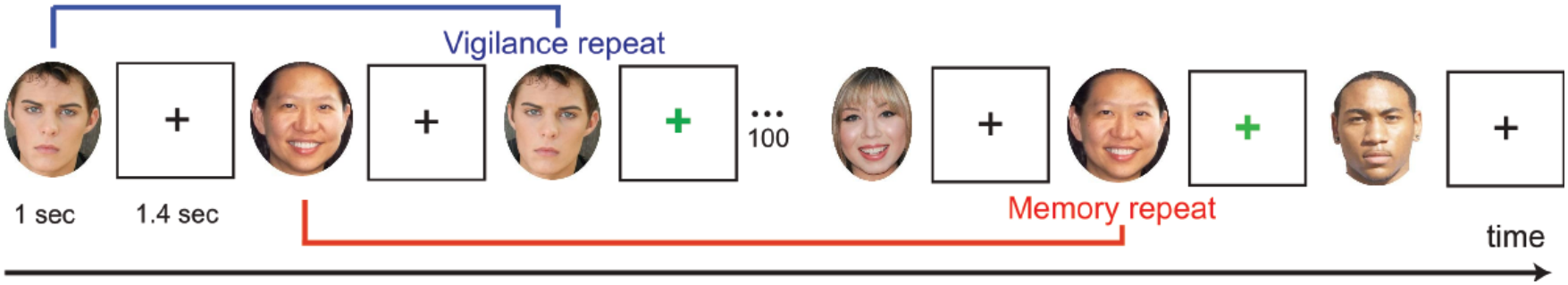
Is memorability consistent across different observers? Yes



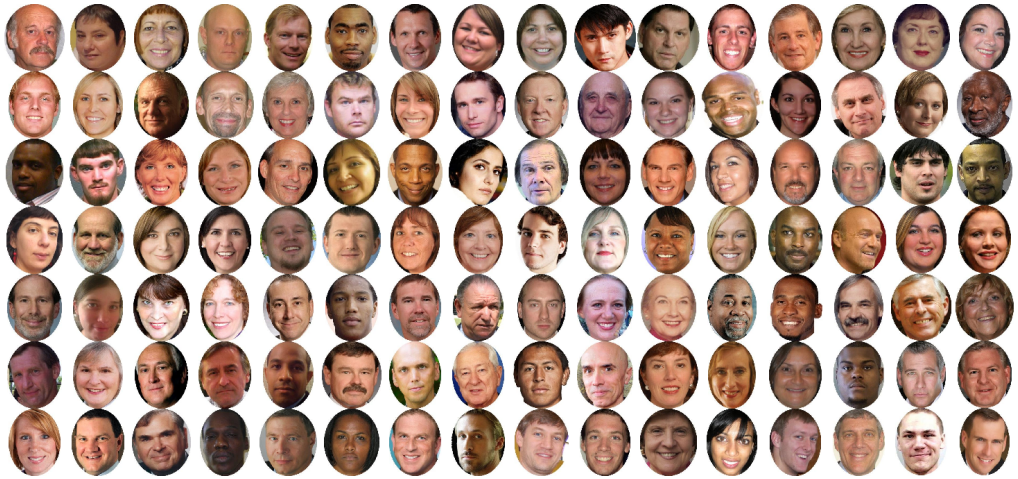
chance level calculated by randomly ordering the images on the x-axis

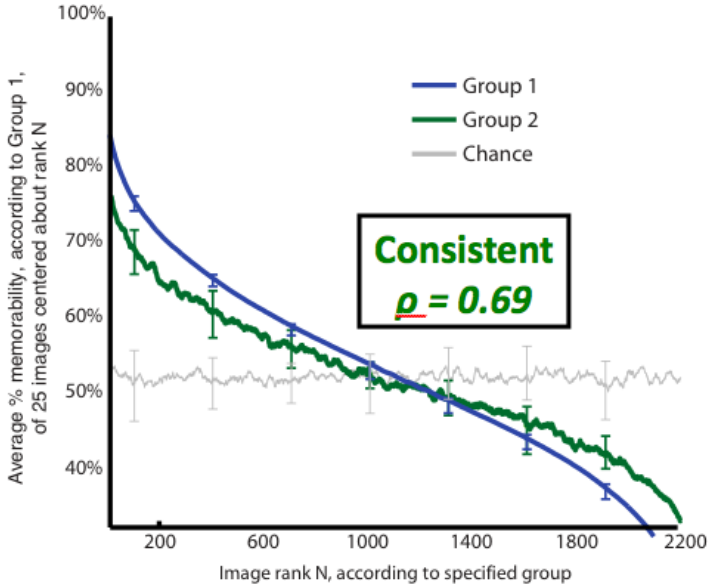
~ 80 scores per image

How consistent is memory within a single stimulus category?



~10,000 unfamiliar faces, 2222 targets with ~ 80 memorability scores

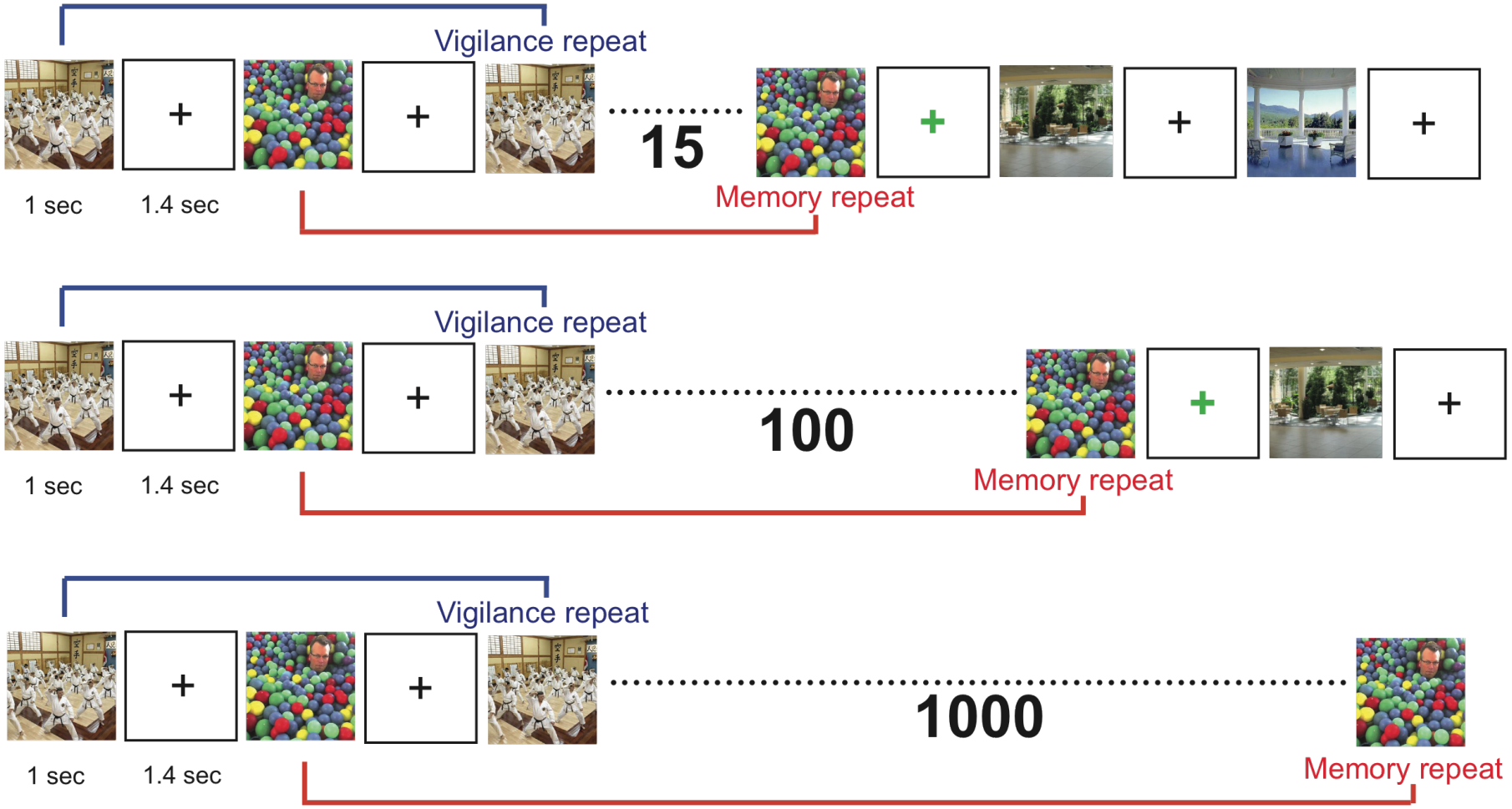


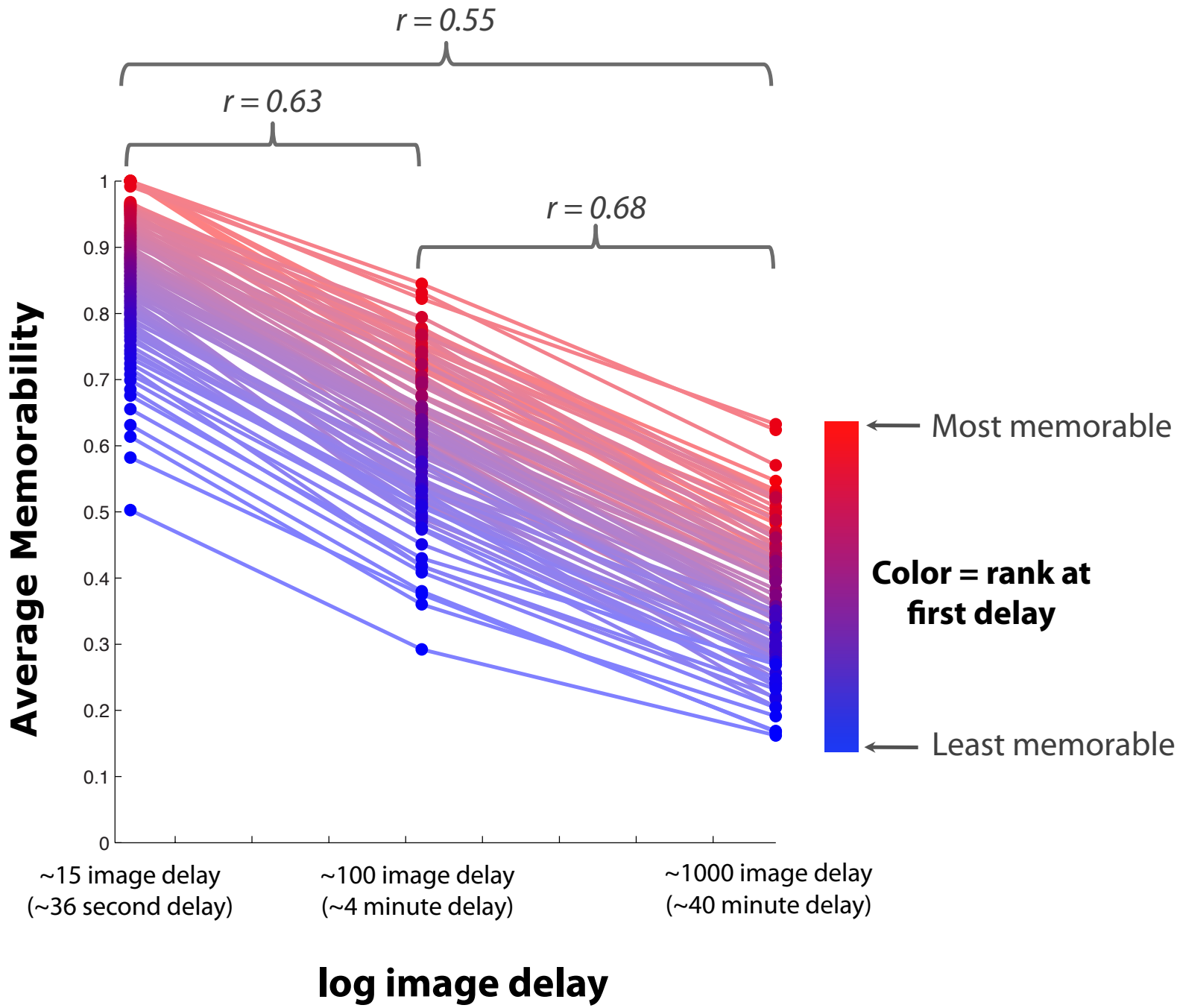


Novel dataset: faces selection based on randomly generated first+last names following the distribution of the US census

Bainbridge, Isola, Blank & Oliva (in press, 2012). Establishing a Database for Studying Human Face Photograph Memory.

Is memorability stable across time?



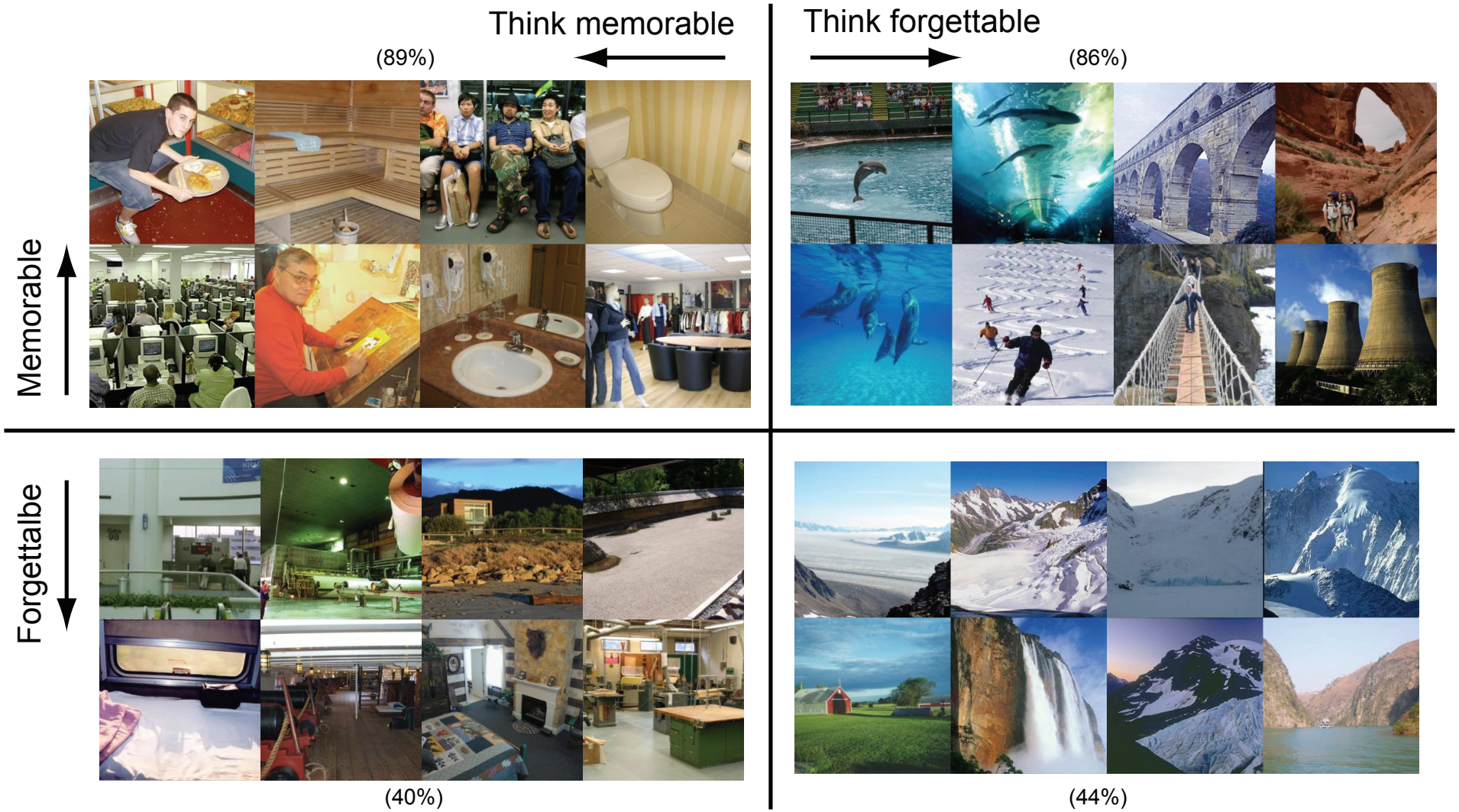

r = 0.55
r = 0.63
r = 0.68
Average Memorability
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
~15 image delay
(~36 second delay)
~100 image delay
(~4 minute delay)
~1000 image delay
(~40 minute delay)
log image delay
Most memorable
Color = rank at first delay
Least memorable

When do memorability differences arise?

At stage of encoding: Some images are encoded in less sufficient detail than others

Memorability rank changes very little over a wide portion of the memory trace!

Subjective judgments do not predict image memorability



Isola et al (2011). *Neural Information Processing Systems (NIPS)*

Can we estimate image memorability?

Humans

0.14 vs **0.89**

0.90 vs **0.44**

Human estimate
True Memorability

Computers

Input Image

Image Features

Support Vector Regression

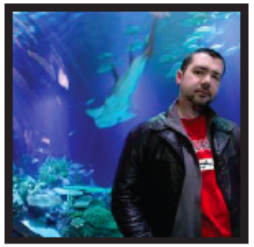
0.73 Memorability Score

Isola et al (2011). *IEEE Proc. Computer Vision & Pattern Recognition (CVPR)*

What makes an image memorable?

- **Intrinsic memorability?**

  Are some images consistently more memorable than others, even across different observers and contexts?

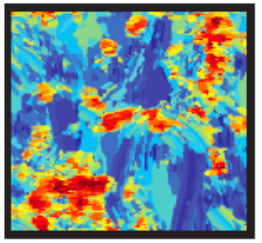


- **What image content matters?**

  What image content (color, object, region) is driving memorability?

- **Can we predict it?**

  Can we automatically predict an image's memorability from its image features?



Prediction algorithm

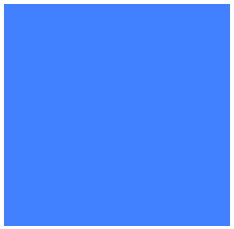
→ **Features** → **Prediction function** → **0.73**

Estimation of image memorability

The result of the regression will be a function that will take as input the features of an image and will output an estimate of the image memorability

Isola et al (2011), CVPR; Isola et al (2011), NIPS

Which features types predict memorability?



1) Simple scalar stats?

brightness, number of objects, mean hue

"Aquarium"

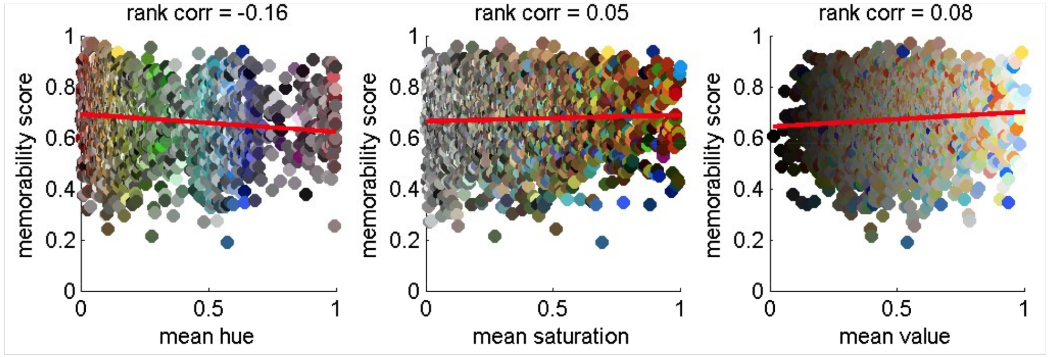
2) Scene category?

e.g. Aquarium, broadleaf forest, art studio

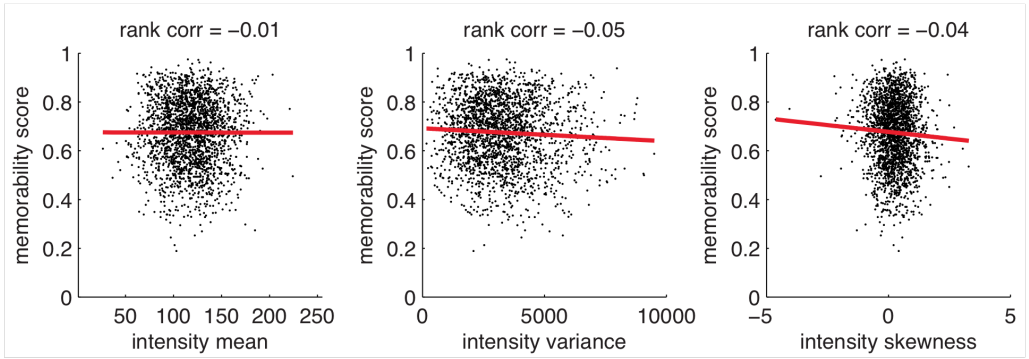
3) Object content?

number, size, and rough position of each object class

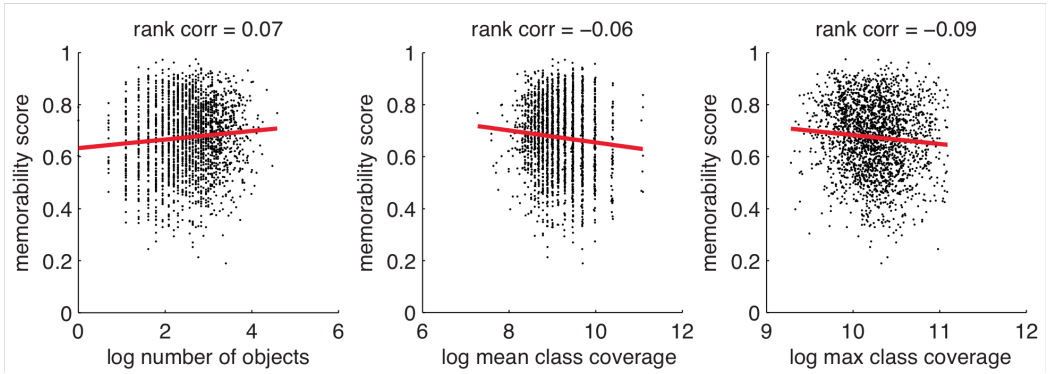
Simple, scalar summary statistics do not correlate well with memorability



Color stats

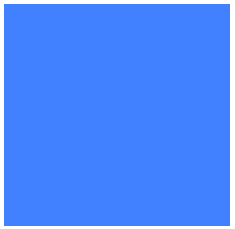


Intensity stats



Object stats

Which features types predict memorability?



1) Simple scalar stats?

brightness, number of objects, mean hue

$\rho < 0.16$

"Aquarium"

2) Scene category?

e.g. Aquarium, broadleaf forest, art studio

3) Object content?

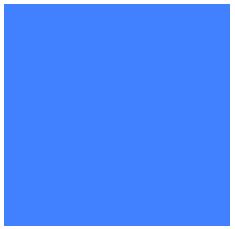
number, size, and rough position of each object class

Scene features

“aquarium”

Categories from Xiao et al. CVPR, 2010

Which features types predict memorability?



1) Simple scalar stats?

brightness, number of objects, mean hue

$\rho < 0.16$

"Aquarium"

2) Scene category?

e.g. Aquarium, broadleaf forest, art studio

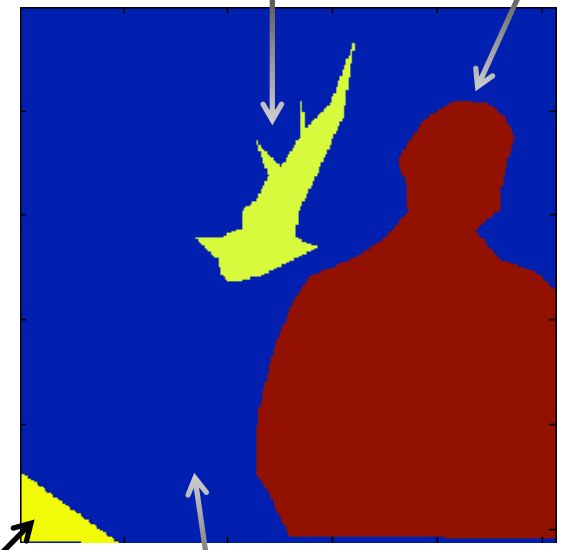
$\rho = 0.37$

3) Object content?

number, size, and rough position of each object class

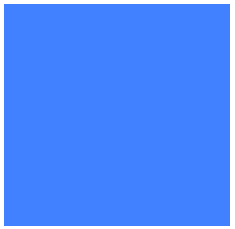
Object features

LabelMe



Segmentations from Choi et al. CVPR, 2010

Which features types predict memorability?



1) Simple scalar stats?

brightness, number of objects, mean hue

$\rho < 0.16$

"Aquarium"

2) Scene category?

e.g. Aquarium, broadleaf forest, art studio

$\rho = 0.37$

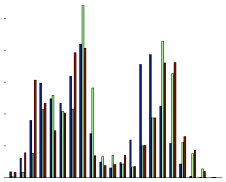
3) Object content?

number, size, and rough position of each object class

$\rho = 0.48$

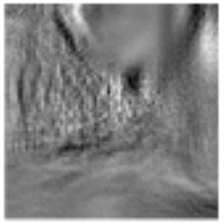
Global image features

Pixel histograms



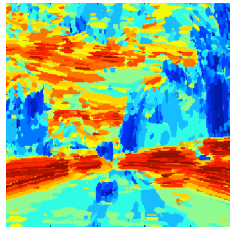
GIST

Gist



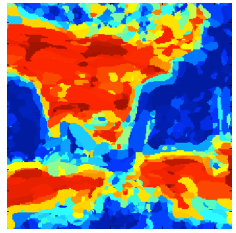
SIFT

Scale-invariant feature transform



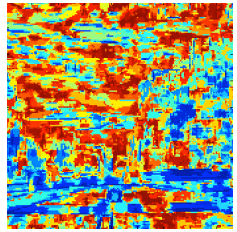
HOG

Histogram of oriented gradients

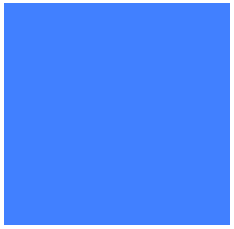


SSIM

Self-similarity



Which features types predict memorability?



1) Simple scalar stats?

brightness, number of objects, mean hue

ρ < 0.16

"Aquarium"

2) Scene category?

e.g. Aquarium, broadleaf forest, art studio

ρ = 0.37

3) Object content?

number, size, and rough position of each object class

ρ = 0.48

4) Global image features?

ρ = 0.46

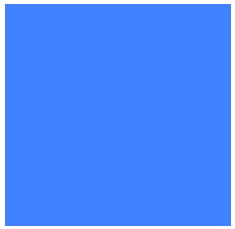
Human consistency: ρ = 0.75

Can we estimate image memorability?

0) Human guessing?

asking people how memorable an image is

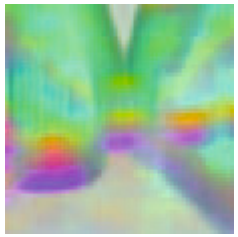
$\rho = -0.02$



1) Simple scalar stats?

color, brightness, number of objects, mean hue

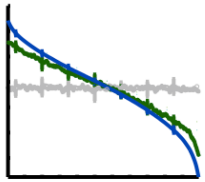
$\rho < 0.16$



2) Computer Model?

SIFT, HOG, GIST and SSIM

$\rho = 0.46$



3) Human objective estimation?

consistency across human subjects in memory game

$\rho = 0.75$

Isola et al (2011). *IEEE Proc. Computer Vision & Pattern Recognition (CVPR)*

Human consistency

$\rho = 0.75$

a) Most memorable images (86%)

b) Typical images (74%)

c) Least memorable images (34%)

Prediction by image features

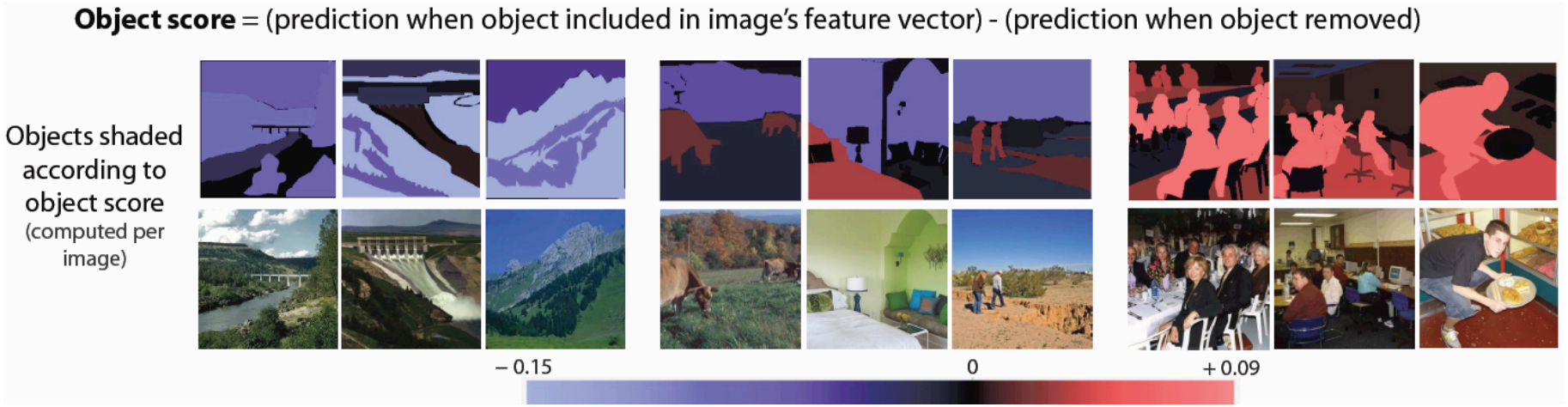
$\rho = 0.46$

a) Predicted most memorable (87%)

b) Predicted typical memorability (68%)

c) Predicted least memorable (52%)

What content makes an image memorable?



Publications

- Bainbridge, W., Isola, P., Blank, I., & Oliva, A. (2012). Establishing a Database for Studying Human Face Photograph Memory. Proceedings of the Cognitive Science Society.

- Isola, P., Xiao, J., Torralba, A., & Oliva, A. (2011). What makes an image memorable? Proceedings of the 24rd IEEE Conference on Computer Vision and Pattern Recognition (pp. 145-152).

- Isola, P., Parikh, D., Torralba, A., & Oliva, A. (2011). Understanding the Intrinsic Memorability of Images. Neural Information Processing Systems (NIPS).

Phillip Isola
Graduate student

Jianxiong Xiao
Graduate student

Wilma Bainbridge
Graduate student

Aditya Khosla
Graduate student

NSF

NATIONAL INSTITUTES OF HEALTH

Google